Journal of Medical Statistics and Informatics
ISSN 2053-7662 | Volume 5 | Article 1

HOAJ
Herbert Open Access Journals

Original Research Open Access

Assessing inter-rater agreement between multiple medical instruments with heteroscedastic measurements

Isuru D. Dassanayake[1,2] and Lakshika S. Nawarathna[2*]

*Correspondence: lakshikas@pdn.ac.lk

[1]Department of Mathematics and Statistics, Texas Tech University, Broadway and Boston, Lubbock, TX 79409-1042, USA.
[2]Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya 20400, Sri Lanka.

Abstract

Background: In clinical medicine, agreement evaluation plays a major role in determining the compatibility and the accuracy of newly introduced methods with pre-existing methods. These methods may be assays, clinical observers, medical devices etc. It is vital to assess the compatibility and the accuracy of these newly introduced techniques because they deal with the measurements of the human body, such as blood pressure, cholesterol level, heart rate etc. In practice, agreement evaluation is carried out among two methods of measurements and deals with the data that are homoscedastic. The main objective of this study is to extend the standard mixed model to allow the error variances to depend on magnitude of measurement and evaluate agreement between multiple methods assuming the new model, taking the heteroscedasticity into account.

Methods: In order to assess the agreement, there are two typical steps in method comparison studies. The first step is to model the data using the Heteroscedastic mixed effects model. The model fitting is carried out by using two main approaches, namely the mean method and the best linear unbiased predictor method. After fitting the model for the second step, the agreement evaluation is carried out using Concordance correlation coefficient and Total deviation index.

Results: The illustrative example contained five methods of measurements and was with heteroscedastic measurements. First, the model fitting was carried out according to the two approaches and the resulting parameters were almost identical. After the model fitting, the agreement evaluation was performed. According to the values resulted from the agreement measurements, it is clear that all five methods agree sufficiently well with the reference method.

Conclusions: The proposed model can be used to model the method comparison data with heteroscedastic measurements with multiple methods of measurements as well as balanced and unbalanced data designs. Under the proposed model, the agreement evaluation methodology for comparing multiple methods is also developed taking heteroscedasticity into account.

Keywords: Agreement evaluation, best linear unbiased predictor, concordance correlation coefficient, heteroscedasticity, mixed effects model, total deviation index

Introduction

In the field of clinical medicine, measurements of the human body take a major part of diagnostic, prognostic and therapeutic evaluations. Due to the rapid advancement technology, new methods and instruments are introduced into this field. These introduced instruments or methods might be more advance, cheaper and easier to use than the old standard instruments. Before these newly introduced methods or instruments put into use, the accuracy and precision of the measurements need to be verified. If these instruments agree sufficiently well with the already existing methods, they can be used interchangeably. To get an understanding of these measurements, they must be compared against with a well-established technique. Then we need to assess the degree of agreement between these methods [1-5].

Experts in this field have already introduced many techniques to test the degree of agreement between two methods. But most of these techniques can be applied under certain assumptions.

doi: 10.7243/2053-7662-5-1

In order to assess the agreement between the assays, the first step is to model the data. The Linear mixed effects model is most commonly used in modeling the method comparison data [6-15]. Because of the flexibility in modeling of within subject dependence, linear mixed models are popular. In this model, normality is assumed for the error terms, but in practice sometimes the normality assumption is violated. In such cases, linear mixed models cannot be used to model the data [16].

To overcome this problem [17] proposed a robust mixed model called "General skew-t mixed model (GSTMM)" that assumes a multivariate skew-t distribution for random effects and an independent multivariate t-distribution for the errors. But this general skew-t mixed model (GSTMM) and most of the other existing models available for method comparison studies are based on the assumption that the variability of the continuous measurement remains a constant throughout the range of the measurement. Though, this is not the case in some practical situations and the variability of the measurements might change with the magnitude, i.e., the 'Heteroscedasticity' of error terms [18].

A novel model to method comparison data with heteroscedastic error variances is proposed in [15], to evaluate the agreement between two methods of measurements measuring continuous data. However, this model cannot accommodate the comparison of multiple methods of heteroscedastic measurements. Therefore, the main objective of this study is to propose a heteroscedastic mixed effects model to analyze heteroscedastic method comparison data with more than two methods of measurements and to adapt the agreement evaluation methodology for multiple methods, taking heteroscedasticity into account.

The lung tumor size measurements data from [19] which motivated this work and are analyzed later in this article, provide a specific example of this phenomenon. The information was gathered from August 2000 to May 2001. In the dataset, out of 33 patients there are 40 lung tumors. These 40 lung tumors belong to 16 men and 17 women from the ages 43 to 78 years. All the lung tumors are larger than 1.5 cm in maximal diameter. Computed Tomography (CT) images of these 40 lung tumors are distributed among five radiologists. All five radiologists are with Thoracic Fellowship training and have more than four years of post-training experience. The five radiologists measured the lung tumors on the CT images by using a ruler or with calipers. Each of these measurements is performed independently and also each of these images was measured twice by each inspector. Here we will consider the measurements taken by the five radiologists as different methods of measurements. Therefore, the dataset contains 40 lung tumors (subjects), 5 readers (methods) and 2 replications for each measurement. In total 40*5*2=400 records.

The rest of this article is organized as follows. Section 2 presents the proposed methodology to deal with heteroscedastic method comparison data measured by multiple methods or raters. Section 3 discusses agreement evaluation under the proposed model using two techniques of model fitting and the last section discusses the results and conclusions of the study. The model fitting and the analysis was carried out by using the R statistical software.

Methodology

In this study, the methodology was divided into two main parts. The first step is to model the data set using an appropriate model and the second step is to measure the agreement among the methods of measurements. The most popular statistical modeling technique in method comparison data is the "Liner mixed effects model". This model is an extension of linear regression models [12] and contains both fixed effects and random effects. In practice the mixed effect models are important when there are repeated measurements. The standard mixed model in the matrix form can be represented as follows.

$$Y_i = X_i\beta + Z_i b_i + e_i; i = 1,\ldots,n, \quad (1)$$

where, Y_i is the vector of observed responses on the i^{th} subject, X_i, Z_i are design matrices for fixed and random effects, β is the vector of fixed effects, b_i is the vector of random effects, e_i is the error vector and e is the vector of all unknown parameters in the model.

The main assumptions of this model are the random effects, random errors have normal distributions and they are mutually independent and the model has a constant error variance which depends only on the method.

$$b_i \sim independent\ N(0,\Psi); e_i \sim independent\ N(0,R_i);$$

$$R_i = diag\left(\sigma_1^2,\ldots,\sigma_1^2,\sigma_2^2,\ldots,\sigma_2^2,\ldots,\sigma_n^2,\ldots,\sigma_n^2\right)$$

where Ψ is a *lxl* positive definite matrix with diagonals Ψ_1^2,....., Ψ_1^2, σ_j^2 is the error variances of method *j*. After fitting the dataset with the mixed effects model, the residuals were analyzed to identify whether there are any model assumption violations.

Heteroscedastic Mixed Effects Model for multiple methods

This model is used when the assumptions (i.e., the random effects and the random errors are normally distributed, independent and has a constant variance) of the standard mixed effect model (homoscedastic model) are violated where error variance changes with the magnitude of the measurements. The heteroscedastic model is as follows.

Let Y_{ijk} is the k^{th} replicate measurement by the j^{th} method on the i^{th} subject, where k=1,…,n_{ij}, j=1,…,l and i=1,…, m. Here n_{ij} is the number of measurement from the j^{th} method on the i^{th} subject, β_j is the fixed mean of the j^{th} method, b_{ij} is the random effect on the i^{th} subject on the j^{th} method, e_{ijk} is the within subject random error, v is the variance covariate and the v_i is values for the i^{th} subject, $z(\rho_{ij})$ is the Fisher's z-trans-

doi: 10.7243/2053-7662-5-1

formation of the correlation coefficient between methods *i* and lastly δ is the heteroscedastic parameter, when δ=0 this assumes homoscedasticity, $\mu_{ij}=\beta_j+b_{ij}$ denotes the conditional mean of $E(Y_{ijk}|b_i)$ and $v_i=h(\mu_i)$ denotes that the v_i is a function of μ_i and $\Sigma_{ij}(v_i)$ is $n_{ij}\times n_{ij}$ diagonal matrix and $\Sigma_{ij}(v_i)$ is $n_i\times n_i$ diagonal matrix where,

$$b_i=(b_{i1},b_{i2},\ldots,b_{il})$$

$$\Sigma_{ij}(v_i)=diag\left\{\sigma_j^2 g_j^2(v_i,\delta_j),\ldots,\sigma_j^2 g_j^2(v_i,\delta_j)\right\}$$

and

$$\Sigma_i(v_i)=diag\left\{\Sigma_{i1}(v_i),\Sigma_{i2}(v_i),\ldots,\Sigma_{ij}(v_i)\right\}$$

Variance covariate is a function of magnitude of measurement μ_i that will be used to model the error variances. A model for conditional error variance is as follows,

$$\text{var}\left(e_{ijk}|b_i\right)=\sigma_j^2 g_j^2(v_i,\delta_j),\mu_i=\beta+b_i,j=1,2,\ldots,l,$$

where g_j is the variance function and δ_j is the heteroscedasticity parameter of the method *j*. According to [12] there exists some common variance functions such as, power model: $(v,\delta)=|v|^{\delta}$, constant plus power model:

$g(v,\delta)=\delta_0+|v|^{\delta_1}$ and exponential model:

$g(v,\delta)=\exp(\delta_v)$. The heteroscedastic model in the matrix form is as follows.

$$Y_i=X_i\beta+Z_i b_i+e_i;i=1,\ldots,m$$

$$e_i|b_i\sim independent\ N_{ni}(0,\Sigma_i(v_i)),\ b_i\sim independent\ N_l(0,\Psi).$$

Due to the scarcity of closed form for the likelihood functions, the exact modeling approaches will be troublesome. Therefore, the model fitting is carried out by two model approximations. First using the mean vector of the reference method as the covariate and the second is using the Best Linear Unbiased Predictor (BLUP) as the covariate. In this case the likelihood functions will be possible in a closed form. Score test and likelihood ratio tests have been carried out to assess the validity of these models. After confirming the models validity, the agreement evaluation is done by using Concordance correlation coefficient (CCC) and Total deviation index (TDI).

Model fitting

The model fitting for this heteroscedastic mixed effects models is carried out by selecting an appropriate value for the variance covariate. In this study, there are two main options for the variance covariate, namely mean of the reference method and the BLUP as the variance covariate. The power model function $g(v,\delta)=|v|^{\delta}$ was selected to fit the model [12].

Using observable mean measurement as variance covariate

As the first approach, we select the mean vector of the reference method as the variance covariate. In this case μ_i^*, an observable quantity is fixed and can fit the model by maximum likelihood method. Here the μ_i^* is expected to be close to μ_i and $\mu_i=\overline{Y}_i$.

Using BLUP as the covariate

As the second approach, we use the best linear unbiased predictors (BLUP) as the variance covariate and fit the data using the heteroscedastic model. Random effects and error terms are independent in this heteroscedastic model and according to the [12] the BLUP can be written as,

$$\mu_{i,blup}=E(b_i|Y_i)=\beta+b_{i,blup};\ b_{i,blup}=E(b_i|y_i);\ v_i^*=h(\mu_i^*).$$

In order to calculate the $\mu_{i,blup}$ any statistical software can be used. v_i^* depends on the unknown parameter θ. The method of calculating the $\mu_{i,blup}$ is an iteratively reweighted method. This method has two steps and it is continued until it converges [12]. More details of this can be found on [15].

When considering the two approaches, the method of using the true mean as the covariate is much simpler than the method of using BLUP as the covariate, because of the complexity of calculating the BLUP than calculating the mean. However, method of using BLUP is more accurate than the other [14].

Agreement evaluation under the proposed Methodology

In health science researches, agreement evaluation is a topic which has considerable interest. This is assessing the agreement between two or more methods measuring the same response [20,21]. In this Section we discuss two measurements that are used to assess the agreement in this study [22,23].

Concordance correlation coefficient (CCC)

Concordance correlation coefficient (CCC) is one of the most common measurements used in order to assess the agreement between methods of measurements. This was introduced by [22]. CCC value ranges between -1 to 1. The higher the values it gives a better agreement. Concordance correlation coefficient under the proposed heteroscedastic mixed model is follows for the lung tumor size measurements,

$$CCC(v_0)=\frac{2\psi_{1j}}{(\beta_1-\beta_j)^2+\psi_1^2+\sigma_1^2 g_1^2(v_0,\delta_1)+\psi_j^2+\sigma_j^2 g_j^2(v_0,\delta_j)};j=2,3,4,5.$$

This represents the concordance correlation coefficient between the reader1 and the reader *j*. For greater accuracy of the measurement CCC was first calculated with Fisher's z-transformation and then converted into the CCC [24].

Total Deviation Index (TDI)

Total deviation index (TDI) is another common measurement of evaluating agreement between two methods of measurements. TDI is the π_0^{th} percentile of absolute value of the differences between the methods (μ_d), for a given large probability π_0. TDI always takes a positive value and the smaller value for TDI indicates good agreement. Under the proposed heteroscedastic model, Total deviation index (TDI) between the reader1 and the reader *j* is defined as follows,

$$\tau^2(v_0)=\psi_1^2+\psi_j^2-2\psi_{1j}+\sigma_1^2 g_1^2(v_0,\delta_1)+\sigma_j^2 g_j^2(v_0,\delta_j);$$

doi: 10.7243/2053-7662-5-1

$$TDI\left(v_0, p_0\right) = \tau\left(v_0\right)\{\chi_1^2\left\{ p_0 \left\{ \frac{\beta_1 - \beta_j}{\tau\left(v_0\right)} \right\}^2 \right\}^{1/2}; j = 2,3,4,5; p_0 = 0.95$$

Results

The initial model fitting was carried out by using the standard mixed effects model. To check the normality of the residuals for each method, the quantile-quintile plot and the Shapiro-Wilk normality test was used. **Table 1** represents the results of the Shapiro-Wilk normality test, which was applied to the residuals of each reader of the mixed effect model. According to the results, it is obvious that the residuals do not follow a normal distribution as all p-values are small. Therefore, the main assumption for the standard mixed effects model (i.e., the residuals have a constant variance) is violated. Moreover, **Figure 1** shows separate quantile-quantile plot for each reader when standard mixed effects is fit to the data. The circles cross the line three times indicates that the hump is not the right shape for these data to be normal. These data are therefore not exactly normal. Hence, we model the error variability using two approximations for μ_i, namely mean of the reference method and the BLUP of μ_i.

Table 1. Shapiro-Wilk test for Normality results.

	Test Statistic	P-value
Reader1	0.96085	0.01516
Reader2	0.93549	0.00056
Reader3	0.97508	0.01197
Reader4	0.95035	0.00361
Reader5	0.95356	0.00554

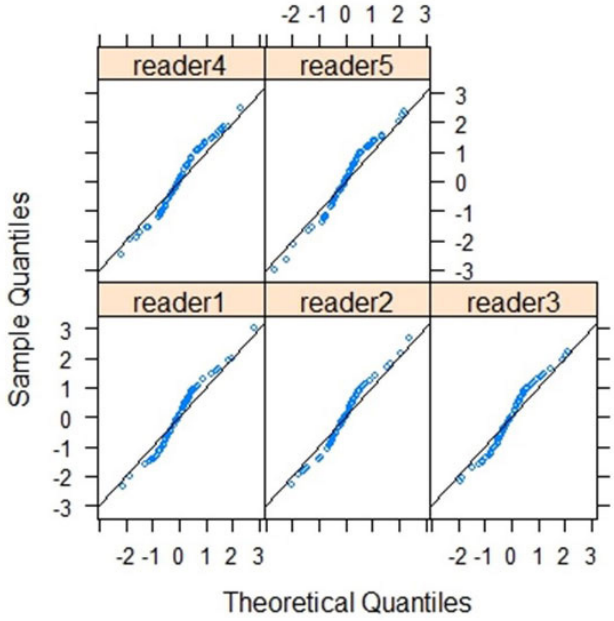

Figure 1. Separate quantile-quantile plot for each reader when the homoscedastic model is fit to the lung tumor size measurements data.

We consider reader1 as the reference method. As the first approach, let's take the mean vector of the reader1 as the variance covariate and fit the data into a heteroscedastic model.

$$\mu_{i1} = \overline{Y}_1$$

In the dataset there are 40 subjects and each subject has 2 replicates. Since there are 5 readers in the dataset, we need to compare 5 different methods of measurements. Therefore, as for the model stated in the methodology, $i=1,2,\ldots,40;\ k=1,2;\ j=1,2,3,4,5.$

The **Table 2** represents the estimates and standard errors of the fitted heteroscedastic model parameters, which was fitted by using mean of the reference method and BLUP as the variance covariate. Considering the results from the **Table 2**, both approaches take approximately equal values when compared with the other methods. Estimator of the reader3 has the largest men value, and the smallest estimator is from reader2.

The Likelihood ratio test has been carried out to identify the most appropriate model among the standard mixed effects model and the proposed heteroscedastic model. The test rejected the null hypothesis, which implies that the reduced model is more appropriate, confirming the heteroscedastic model is the best among these two models. Furthermore, the score test is also performed on these models. The result was the same as before confirming that the heteroscedastic model is more suitable for the given dataset. Since the fitted heteroscedastic model is appropriate the next step is to assess the agreement between the methods of measurements.

Agreement evaluation under the Heteroscedastic model

The **Figure 2** represents the estimates and their confidence intervals of CCC and TDI using the mean method. The solid

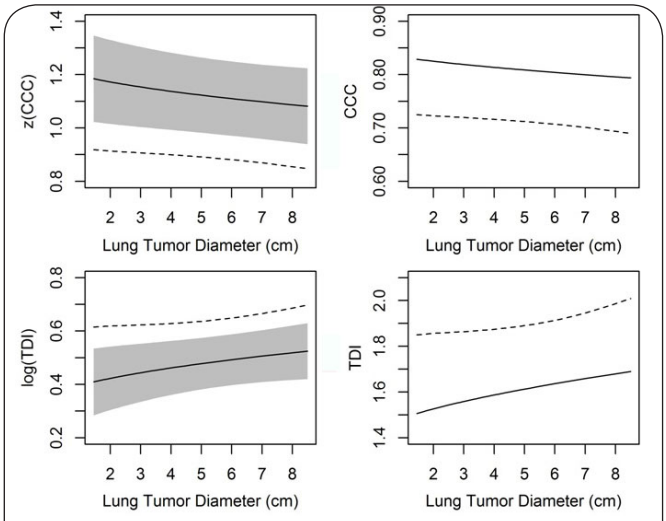

Figure 2. Estimates and their 95% CI of CCC and TDI between reader1 and reader2 using mean method.

doi: 10.7243/2053-7662-5-1

Table 2. Model parameter estimates and their standard errors using both mean and BLUP based methods.

	Mean method		BLUP method	
	Estimate	Standard Error	Estimate	Standard Error
β_1	4.0218	0.2547	4.0237	0.2544
β_2	3.5947	0.2297	3.5941	0.2296
β_3	4.3361	0.2305	4.3364	0.2307
β_4	4.2703	0.2549	4.2696	0.2551
β_5	4.1023	0.2536	4.1049	0.2536
$\log\sigma_1^2$	-2.8360	0.8805	-2.6970	0.8440
$\log\sigma_2^2$	-3.2381	0.9919	-3.5125	1.1379
$\log\sigma_3^2$	-3.9337	0.6523	-3.9942	0.6966
$\log\sigma_4^2$	-3.0002	0.7641	-3.0398	0.8109
$\log\sigma_5^2$	-0.5715	0.6754	-0.5420	0.7226
$\log\psi_1^2$	0.9352	0.2280	0.9330	0.2280
$\log\psi_2^2$	0.7188	0.2312	0.7166	0.2320
$\log\psi_3^2$	0.7334	0.2300	0.7342	0.2301
$\log\ _4$	0.9305	0.2304	0.9310	0.2305
$\log\psi_5^2$	0.9278	0.2267	0.9266	0.2268
$z(\rho_{12})$	1.4140	0.1892	1.4265	0.1962
$z(\rho_{13})$	1.3964	0.1723	1.4017	0.1727
$z(\rho_{14})$	1.9112	0.2033	1.9288	0.2060
$z(\rho_{15})$	1.8785	0.2032	1.8806	0.2014
$z(\rho_{23})$	1.1959	0.1781	1.2048	0.1817
$z(\rho_{24})$	1.3909	0.1908	1.4061	0.1981
$z(\rho_{25})$	1.2123	0.1771	1.2194	0.1816
$z(\rho_{34})$	1.6048	0.1870	1.6009	0.1863
$z(\rho_{35})$	1.6531	0.1835	1.6601	0.1843
$z(\rho_{45})$	2.0998	0.2299	2.1073	0.2322
δ_1	0.1622	0.3028	0.1150	0.2897
δ_2	0.4010	0.3986	0.5087	0.4618
δ_3	0.5385	0.2374	0.5608	0.2542
δ_4	0.3500	0.2767	0.3652	0.2944
δ_5	-0.7732	0.2625	-0.7679	0.2814

line indicates the estimate and the shaded region indicates the estimate±standard error. The dotted line represents the lower bound of the concordance correlation coefficient and the upper bound of the Total deviation index. The upper two graphs represent the CCC with Fisher's z transformation and estimated CCC and lower two graphs represent the TDI with log transformation and estimated TDI values. The CCC gradually decreases from 0.82 to 0.79 through the diameter of 1.45 to 8.5 cm. The TDI value increases from 1.505 to 1.69 with the increase of the magnitude of the measurement. From both CCC and TDI values we can observe that for small diameter values the both methods (readers) have satisfactory agreement, while with the increase of the magnitude of the lung tumor diameter, the agreement decreases. Therefore, it is safe to conclude that the reader2 sufficiently agrees with the reference method (reader1). Likewise, the readings from the readers 3, 4 and 5 derived almost similar results as the reader2 by confirming that all the methods (readers) agree well with the reference method. i.e., reader1.

The **Figure 3** represents the CCC and TDI values for the reader1 and reader2, using the BLUP based method. The **Figure 2** and the **Figure 3** are almost identical. Therefore, reader2 is well agreed with the reader1.

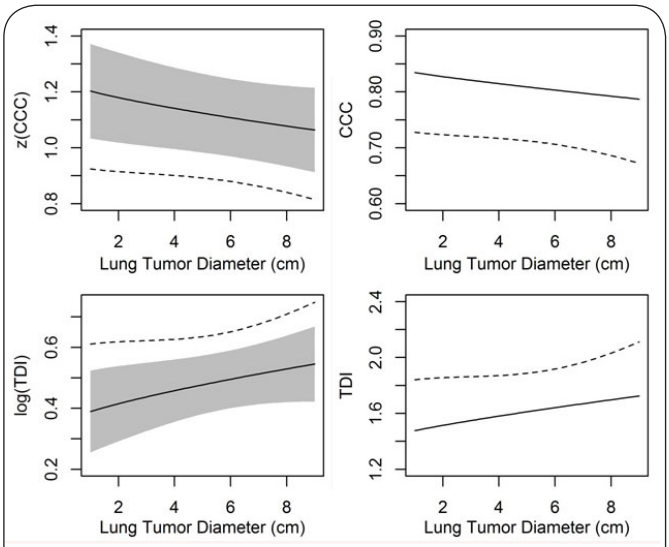

Figure 3. Estimates and their 95% CI of CCC and TDI between reader1 and reader2 using BLUP method.

Discussion

The proposed heteroscedastic mixed effects model was fitted by using two methods. One was by using true mean as the variance covariate and the other method is by using the Best Linear Unbiased Predictor (BLUP) as the variance covariate. When fitting the model by using both methods, it was confirmed that the heteroscedastic model is better than the standard mixed effect model. The results from the likelihood ratio test and the score test confirms that conclusion. When considering the means from the **Table 2**, we can observe that the means from both the approaches are identical if round up to 2 decimal places. This means that the mean method and the BLUP method generate almost identical estimates.

CCC and TDI were used in evaluating the agreement between the reference method (reader1) and the other 4 methods (readers). Both mean based method and the BLUP based method produced almost identical (with slight differences) results for CCC and for TDI. **Figures 2** and **3** both show that reader2 agrees well with reader1. Not only the reader2, but also reader3, reader4 and reader5 are also agreed well with the reader1. Moreover, **Figures 2** and **3** show that the CCC

and TDI values varies with the magnitude. If the data set was modeled using a homoscedastic model, constant values for CCC (0.8069) and TDI (1.6140) can be obtained. Therefore, if a dataset with heteroscedasticity is modeled with standard mixed effects model, the resulting outcome will be misleading and inaccurate. To overcome this issue, the data must be modeled with a model that accounts the heteroscedasticity.

Conclusions

We propose a method successfully in order to fit 5 methods of heteroscedastic clinical measurements and this proposed model can be easily extended to deal with any number of multiple methods which has replicated data. Our approach can accommodate balanced or unbalanced data designs and it works well with any scalar measure of agreement. Moreover, the two model fitting methods discussed give almost identical estimates to the model. The limitation of this study is that the proposed method can be applied only with replicated measurements.

List of abbreviations

BLUP: Best Linear Unbiased Predictor
CCC: Concordance Correlation Coefficient
TDI: Total Deviation Index

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Authors' contributions	IDD	LSN
Research concept and design	✓	✓
Collection and/or assembly of data	✓	✓
Data analysis and interpretation	✓	✓
Writing the article	✓	✓
Critical revision of the article	✓	✓
Final approval of article	✓	✓
Statistical analysis	✓	✓

Acknowledgement

The authors are grateful to Professor L. Broemeling for providing the lung tumor size measurements dataset.

Publication history

Editor: Jimmy Efird, East Carolina University, USA.
Received: 18-Oct-2016 Final Revised: 26-Dec-2016
Accepted: 09-Jan-2017 Published: 27-Jan-2017

References

1. Altman DG and Bland JM. **Measurements in Medicine: The Analysis of Method Comparison studies**. *Journal of the Royal Statistical Society*. 1983; **32**:307-317. | Pdf
2. Bland JM and Altman DG. **Statistical methods for assessing agreement between two methods of clinical measurement**. *Lancet*. 1986; **1**:307-10. | Article | PubMed
3. Bland JM and Altman DG. **Measuring agreement in method comparison studies**. *Stat Methods Med Res*. 1999; **8**:135-60. | Article | PubMed
4. Choudhary PK. **Interrater Agreement**. *In methods and applications of statistics in the life and health sciences*, John Wiley: New York. 2009; 461-480. | Book
5. Haber M, Gao J and Barnhart HX. **Evaluation of Agreement between Measurement Methods from Data with Matched Repeated Measurements via the Coefficient of Individual Agreement**. *J Data Sci*. 2010; **8**:457-469. | PubMed Abstract | PubMed FullText
6. Altman DG and Bland JM. **Agreement between methods of measurement with multiple observations per individual**. *Journal of biopharmaceutical statistics*. 2007; **17**:571-582.
7. Bland JM and Altman DG. **Agreement between methods of measurement with multiple observations per individual**. *J Biopharm Stat*. 2007; **17**:571-82. | Article | PubMed
8. Barnhart HX, Haber M and Song J. **Overall concordance correlation coefficient for evaluating agreement among multiple observers**. *Biometrics*. 2002; **58**:1020-7. | Article | PubMed
9. Carrasco JL and Jover L. **Estimating the generalized concordance correlation coefficient through variance components**. *Biometrics*. 2003; **59**:849-58. | Article | PubMed
10. Carrasco JL, King TS and Chinchilli VM. **The concordance correlation coefficient for repeated measures estimated by variance components**. *J Biopharm Stat*. 2009; **19**:90-105. | Article | PubMed
11. Carstensen B, Simpson J and Gurrin LC. **Statistical models for assessing agreement in method comparison studies with replicate measurements**. *Int J Biostat*. 2008; **4**. | PubMed
12. Pinheiro JC and Bates DM. **Mixed-Effects Models in S and S-PLUS**. *Springer: New York*. 2000. | Book
13. Roy A. **An application of linear mixed effects model to assess the agreement between two methods with replicated observations**. *J Biopharm Stat*. 2009; **19**:150-73. | Article | PubMed
14. Nawarathna LS and Choudhary PK. **A heteroscedastic measurement error model for method comparison data with replicate measurements**. *Stat Med*. 2015; **34**:1242-58. | Article | PubMed
15. Nawarathna LS. **Heteroscedastic Models for Method Comparison Data**. (Doctoral Dissertation). ProQuest Dissertation and Theses Database. 2014. | Article
16. Nawarathna LS and Choudhary PK. **Measuring agreement in method comparison studies with heteroscedastic measurements**. *Stat Med*. 2013; **32**:5156-71. | Article | PubMed
17. Choudhary PK, Sengupta D and Cassey P. **A general skew-t mixed model that allows different degrees of freedom for random effects and error distributions**. *Journal of Statistical Planning and Inference*. 2014; **147**:235-247. | Article
18. Hawkins DM. **Diagnostics for conformity of paired quantitative measurements**. *Stat Med*. 2002; **21**:1913-35. | Article | PubMed
19. Erasmus JJ, Gladish GW, Broemeling L, Sabloff BS, Truong MT, Herbst RS and Munden RF. **Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response**. *J Clin Oncol*. 2003; **21**:2574-82. | Article | PubMed
20. Choudhary PK and Yin K. **Bayesian and Frequentist methodologies for analyzing method comparison studies with multiplemethods**. *Statistics in Biopharmaceutical Research*. 2010; **2**:122-132. | Article
21. Dunn G and Roberts C. **Modelling method comparison data**. *Stat Methods Med Res*. 1999; **8**:161-79. | PubMed
22. Lin LI. **A concordance correlation coefficient to evaluate reproducibility**. *Biometrics*. 1989; **45**:255-68. | Article | PubMed
23. Lin LI. **Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence**. *Stat Med*. 2000; **19**:255-70. | Article | PubMed
24. Lin LI, Hedayat AS, Sinha B and Yang M. **Statistical methods in assessing agreement: models, issues, and tools**. *Journal of the American Statistical Association*. 2002; **97**:257-270. | Article

Citation:
Dassanayake ID and Nawarathna LS. **Assessing inter-rater agreement between multiple medical instruments with heteroscedastic measurements.** *J Med Stat Inform*. 2017; **5**:1. http://dx.doi.org/10.7243/2053-7662-5-1